Psychological Review
2005, Vol. 112, No. 3, 694–707

0033-295X/05/$12.00 DOI: 10.1037/0033-295X.112.3.694

Constraints and Nonconstraints in Causal Learning: Reply to White (2005) and to Luhmann and Ahn (2005)

Patricia W. Cheng
University of California, Los Angeles

Laura R. Novick
Vanderbilt University

The task of causal learning concerns figuring out the laws that govern how the world works. The goal of a reasoner who engages in this task is to gain an understanding of the empirical world that would guide decisions regarding actions to achieve the reasoner's objectives. The comments by P. A. White (2005) and C. Luhmann and W.-k. Ahn (2005) on P. W. Cheng (1997) and L. R. Novick and P. W. Cheng's (2004) power PC theory of causal learning do not define the constraints of the task of causal learning in the same way as does that theory: They change constraints on the input and omit consideration of the goal. This article clarifies the approach taken by the power PC theory to address the issues raised. In particular, it illustrates how the approach provides a framework for answering causal questions under various assumptions—a framework that allows the incremental construction of a causal picture of a complex world.

Keywords: causal learning, discovery, attribution, hypothesis testing, problem solving

The comments by White (2005) and Luhmann and Ahn (2005) criticize various aspects of our power PC theory (Cheng, 1997, 2000; Novick & Cheng, 2004), which concerns how a reasoner comes to know that one thing causes another. We welcome this opportunity to address the issues raised. White's criticisms, which we address first, provide a springboard for us to more clearly illustrate a major appeal of our approach—the ability to explain a variety of causal reasoning phenomena in a logically consistent manner. We show that the various pieces of evidence White offers as refutations of the power PC theory (with the exception of findings that appear to be due to confounding) are actually consistent with our approach. Luhmann and Ahn's criticisms allow us to discuss several distinctions previously left implicit in the causal learning literature. We likewise explain their criticisms of apparent contradictions in our theory. In summary, we show that our approach allows one to see order where our commentators see chaos.

It may help at the outset to remind readers that the power PC theory, which stands for the causal power theory of the probabilistic contrast model, inherits the probabilistic contrast model (Cheng & Novick, 1990, 1991, 1992) and its explanations of reasoning phenomena (see Cheng, 1997). Also, we note that although our work on causal learning has dealt only with causes and effects that are represented as binary variables, we do not restrict *covariation*—the concept of concomitant changes across variables—to its binary interpretation; the concept applies to other types of variables as well.

Patricia W. Cheng, Department of Psychology, University of California, Los Angeles; Laura R. Novick, Department of Psychology and Human Development, Vanderbilt University.

The preparation of this article was supported by National Institutes of Health Grant MH64810. We thank Woo-kyoung Ahn, Christina Ford, Keith Holyoak, Clark Glymour, and Peter White for their helpful comments on a draft.

Correspondence concerning this article should be addressed to Patricia W. Cheng, Department of Psychology, Franz Hall, University of California, Los Angeles, CA 90095-1563, or to Laura R. Novick, Department of Psychology and Human Development, Vanderbilt University, Peabody #512, 230 Appleton Place, Nashville, TN 37203-5721. E-mail: cheng@lifesci.ucla.edu or laura.novick@vanderbilt.edu

Causal Roles

White (2005) argues that the causal roles in his *causal powers* theory (causal power, liability, and releasing condition), which is based on the philosophical theory of powerful particulars (Harré & Madden, 1975), are incompatible with our theory. He claims that "the most the power PC theory could provide would be an estimate of the conjunctive causal probability of the interaction between the three things—power, liability, and releasing condition" (p. 679). He concludes, "the truth about causality lies in causal role constructions, not in regularities" (pp. 679). In this section, we show how our theory contributes to the assessment of causal roles, and we identify the additional information required for assigning these roles.

Causal Learning as Problem Solving: Causal Roles as Outputs

The task of causal learning can be regarded as a problem to be solved. Problems are defined by the givens and the desired unknown(s). For causal learning, the desired unknowns are various kinds of causal judgments (e.g., causal strength, causal attribution, confidence that a causal relation exists)—the outputs of the inference process. The givens are (a) observations (including introspective sensations) that serve as (noncausal) input to the process and (b) intrinsic assumptions, if any, that mediate the transforming of the input into the output. The critical question with respect to causal learning, as with problem solving more generally, concerns the nature of the process that maps the givens onto the desired unknowns.

We examine White's (2005) causal roles from this perspective. To help clarify the discussion, we introduce two terms. We call the thing in which the effect occurs the *patient,* and we refer to the other things in which a candidate cause occurs as *agents* (the patient also could harbor a candidate cause). In White's example in which the breaking of a plate is the effect in question, the plate is the patient. The fragility of the plate is a candidate cause that occurs in the patient. The act of dropping the plate is a candidate cause that occurs in the toddler, an agent. The hardness of the floor is a candidate cause that occurs in the floor, another agent. We can now translate White's three causal roles in terms of our theory. They are measurable (i.e., observable) properties of agents and patients that are component variables in a conjunctive cause: Causal power is a component that is an enduring property of an agent (in this case, the hardness of the floor); liability is a component that is an enduring property of the patient (the fragility of the plate); and releasing conditions are components that are transient properties of the patient, an agent, or their combination (the dropping of the plate). Needless to say, not all properties of the floor (e.g., its smoothness) have the causal power to break the plate; nor do all properties of the plate (e.g., its color) make it liable to breaking.

Now, one can observe the patient and agents (e.g., plate, toddler) in which the components of a conjunctive cause (the fragility of the plate, the act of dropping) occur as well as whether these components are transient or enduring properties of their respective entities. In fact, White (2005) himself suggests these observable inputs and their relations to the roles. Such information should therefore be legitimate as input to any theory of causal learning, just as it is available to White, to arrive at causal roles as outputs.

In summary, when causal learning is viewed as a problem to be solved, there is nothing incompatible between our theory and White's (2005) causal roles. The measures in our theory provide the criteria for selecting properties that fill these roles (e.g., the fragility rather than the color of the plate) with respect to an outcome (breaking), differentiating them from merely covariational and noncovariational properties.

Evaluation of Liabilities and Releasing Conditions: Candidate Causes as Variables

White (2005) contends, however, that the measures in our theory are incapable of evaluating either releasing conditions or liabilities. Illustrating his point, he writes, "Suppose we have a drug that might produce a certain side effect in patients and we want to ascertain whether this side effect is caused by the means of administration" (p. 680). He objects that to accomplish this task, "We cannot plausibly compare ingesting a pill with not administering the drug at all: That would be a way of testing the causal power of the drug" (p. 680), rather than testing ingesting as a releasing condition.

Contrary to White's (2005) argument, when candidate causes, including releasing conditions and liabilities, are defined as variables (i.e., features, such as the medicine in the pill) rather than as objects (e.g., the pill), the power PC theory explains the evaluation of all three causal roles. Using White's example, we show that comparing methods of administration, as White suggests doing, does not involve a different process of causal inference; it merely allows the same process of probabilistic contrast to be applied to data that better control for alternative causes.

To evaluate a conjunctive cause, as for a simple cause, one needs to avoid confounding (the power PC theory explains this need). The medicine is in the pill—the agent—but it is only one of the candidate causes involving the agent. As we all know, taking a pill may have a placebo effect. Thus, as White (2005) rightly objects, one would not compare ingesting a pill containing the medicine with not administering any pill to test ingestion as a releasing condition; nor would this comparison be a good test of the causal power of the medicine. Such a test would be confounded (medicine in the bloodstream, ingestion, and the cause of the placebo effect would all covary). Instead, one should hold constant the presence of a pill and independently vary all candidate-cause variables of interest. Let us start with varying ingestion and medicine, as shown in Figure 1.[1] The contrast between ingesting a placebo and leaving it in the bottle (the right column in the figure) would indicate whether ingestion per se produces the side effect (Cheng, 1997, discussed a similar example; for one to evaluate a scratch on the skin, a means of administrating a food-allergy test, as the cause of hives [an outcome], there needs to be a scratch by itself to avoid confounding with food, the target candidate cause).

There are other candidate causes to be considered. Consider the outcome patterns shown in Figure 2, in which the side effect is either present or absent (represented by "+" and "o," respectively). Pattern A indicates that ingestion itself produces the side effect (it is a simple power and therefore not a releasing condition in White's, 2005, terms), whereas Pattern B indicates that ingestion and the medicine conjunctively produce the side effect (ingestion is therefore potentially a releasing condition). The evaluation based on Pattern B, however, is coarse grained; one does not know whether (a) the interaction between ingestion and medicine is indeed the cause or (b) merely having the medicine in the bloodstream, an effect of the interaction (a candidate cause downstream from those represented in the figure) rather than the interaction itself, produces the side effect. In the latter case, ingestion still would not be a releasing condition for the side effect in question. This is where White's (2005) intuition comes in: Comparing methods of administration, as he advocates, provides the prerequisite information for controlling for alternative causes, medicine in the bloodstream in this case. Suppose that injecting the medicine shows no side effect, as illustrated by Pattern C. In that case, both medicine in the bloodstream and injection can be ruled out as causes of the side effect. Now, comparing the two methods of administration (Patterns B and C), one can draw the finer grained conclusion that the interaction between ingestion and medicine itself, rather than medicine in the bloodstream, is a cause: Medicine in the bloodstream is held constant across the two patterns. For this conjunctive cause, medicine—being the component that is an observable enduring property associated with the agent—is therefore a causal power, and ingestion—being a com-

[1] Part of the evaluation trespasses on well-learned knowledge (e.g., if a medicine is left in the bottle, it cannot possibly have an effect); readers must therefore suspend their intuition and pretend to acquire that knowledge anew.

ponent that is a transient condition associated with the combination of the agent and the patient—is therefore a releasing condition.[2]

Sometimes, the relevant information may be unavailable or physically impossible; for example, it is impossible to have medicine magically appear in the bloodstream without any method of introducing it. If injection and ingestion both show Pattern B and there is no information on other methods of administration, it would be impossible to tell whether (a) medicine in the bloodstream per se produces the side effect or (b) ingestion and injection each interacts with the medicine to produce the side effect. No theory of causal learning can yield an answer to whether the method of administration is a releasing condition for the side effect in that situation, and they should not. The shortcoming lies in the availability of relevant information (i.e., in the input) rather than in the reasoning process.[3]

In summary, the inference process in our theory can readily be extended to explain White's (2005) causal roles. Specifically, when candidate causes are represented as variables, as they should be, liabilities and releasing conditions are evaluated just like causes: They are components of a conjunctive cause and are further assigned roles by their observable relation to agents and patients. The evaluation of causal roles is therefore fully compatible with the power PC theory.

Abstract Causal Knowledge and the Confusion Between Input and Output

Distinguishing between a singularist and a regularity account of causal reasoning, White (2005) writes,

> The operation of a causal power is not a matter of probability or frequency. . . . In fact a power might never be exercised at all. . . . People can ascribe a causal power to a thing on the basis of knowledge of its nature even if they have no evidence for the operation of the power in question." (p. 677)

Three confusions concerning the problem of causal learning are involved here.

First, White (2005) allows abstract prior causal knowledge in the input. Ascribing causal power from "knowing the nature" of a thing simply means that one is applying abstract prior causal knowledge. For the problem of causal learning, assuming that the input already contains the desired unknown—the domain-specific causal knowledge—would create a circular argument (e.g., Ahn, Kalish, Medin, & Gelman, 1995; Shultz, 1982; White, 2000). It seems that White is addressing a different research problem.

Second, the fact that reasoners use abstract prior causal knowledge does not mean that regularity information is unnecessary for causal learning. One cannot know, for example, that the wind

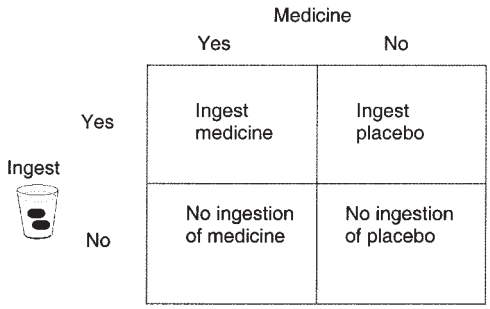

Figure 1. Testing ingestion as a releasing condition while holding "pill" constantly present.

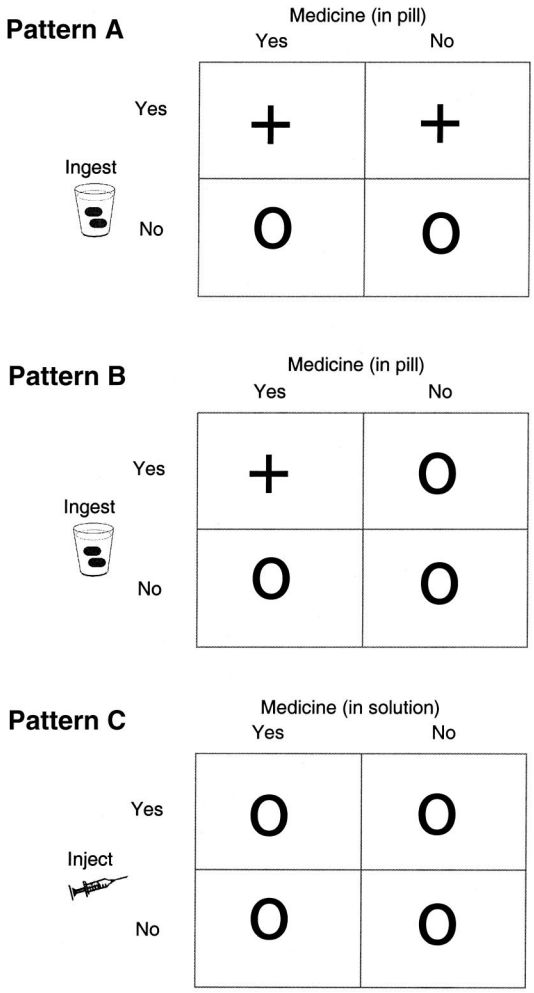

Figure 2. Outcome patterns for different methods of administering medicine holding the agent (pill or solution) constantly present. + = side effect; o = no side effect.

[2] The comparison between methods of administration suggested by White (2005) might be interpreted to mean that only the left column in Figure 2 is relevant to the assessment of a releasing condition and that our theory includes irrelevant information. It should be clear, however, that the right column is relevant: Without that column, one would be unable to tell whether ingestion is a simple cause of the side effect (Pattern A), noncausal by itself (Pattern B), or a releasing condition for the medicine to produce the side effect (Patterns B and C together).

[3] To determine whether one conjunctive cause has more power than another (e.g., whether swallowing a medicine is more effective than injecting it), one would measure the power of each conjunctive cause separately using the standard power PC inference process and then compare the outputs. It is clear that such comparisons do not imply a process that contradicts the standard one.

blows leaves off trees without any relevant force and dislodgment experience. The fact that one might not need to have ever seen the wind blow a leaf off a tree to surmise that it has the power to do so merely means that the relevant causal relation was learned at a more abstract level; for example, a strong force may dislodge an object that is attached to another object. One might have learned that a toddler's hand does not dislodge the snap-on cap from a ketchup bottle but a tidal wave dislodges a boat tied to the dock. Consistent with the literature on analogy (e.g., Holyoak, 2005), one might expect that reasoners learn specific causal relations first (e.g., a tidal wave dislodges a boat, an adult's hand dislodges the cap from the ketchup bottle, the wind blows the cover off a garbage can) and then form abstractions over those cases.

Finally, White's (2005) distinction between *singularist* and *regularity* accounts reveals a confusion between the input and the output for causal learning. The analysis of causal roles, which is the goal of White's causal powers theory, addresses questions about the output: Is there a domain-general structure to causal knowledge, including knowledge regarding a single instance, and if so, what is that structure (e.g., what is one's understanding of the dislodging of one particular leaf from one particular tree)? Recall that regularity theories, which are theories of learning, address a different question: the transformation of the input into the output. Although the output—the resulting causal knowledge—can be applied to a single instance, more than a single instance is required as input; covariation information is necessary, and one needs at least two points on a scatter plot, for example, to see a correlation. Thus, evidence for causal understanding regarding a single instance does not refute theories of causal learning.

Evaluating Empirical Evidence Reviewed by White (2005)

White (2005) reviews three lines of empirical research that he argues provide evidence against the regularity approach in general and our approach in particular. Below we show how each in fact is consistent with our approach.

Generativity Versus Regularity

White (2005) argues that generativity, being the watershed feature, would allow a test between his approach and ours. He writes, "If the causal relation is understood as generative and information about a generative relation is available, then under the causal powers theory, causal inference should be guided by that information, even when it conflicts with regularity information" (p. 678).

The incorporation of generativity in a regularity theory. Let us examine White's (2005) contest between generativity and regularity from our perspective. We agree with White that purely covariational models cannot arrive at (domain-specific) generativity in the output (i.e., a causal belief about a particular domain). The power PC approach solves this problem by incorporating generativity as a *domain-free* assumption in the givens (i.e., as a general conviction that there are such things in the world as causes that produce effects). The incorporation of generativity differentiates the power PC approach from the purely covariational approach that is the actual target of White's objections.

To illustrate how the generativity assumption in our theory explains the differentiation between covariation and causation, we consider two empirical phenomena to which our theory has not previously been applied. Both phenomena concern the insufficiency of covariation for inferring causation (these examples contrast with attempts to show that covariation is not necessary for causal inference, which as we explain is futile).

One phenomenon concerns the situation in which candidate cause c is not a cause of effect e (i.e., has 0 power in our terms) and yet covaries with e. Consider the covariation between a drop in the barometric reading and the approach of a storm, despite the fact that the drop does not cause storms. Below, to illustrate the decoupling of the value of covariation from the value of generative power, we apply Cheng's (1997) Equation 5 to an approximate representation of the barometer reading and storm relation (approximate because the equation does not apply to continuous variables).

Let us represent the barometric reading as either dropping or not dropping, with r representing "a drop in the reading" and $\bar{r}$ representing "no drop in the reading"; likewise, let us represent a storm the next day as either present or not. Let $q_{\text{reading} \rightarrow \text{storm}}$ represent the power of a drop in the barometric reading to cause a storm and $q_{\text{atmosphere} \rightarrow \text{storm}}$ represent the power of a drop in atmospheric pressure to cause a storm.[4] These causal power variables reflect instantiations of the generativity assumption in our theory for the particulars of this situation. The equation would then be instantiated as follows:

$$\begin{aligned}\Delta P_r &= P(\text{storm}|\text{r}) - \text{P}(\text{storm}|\bar{r}) \\ &= q_{\text{reading}\rightarrow\text{storm}} \cdot [1 - P(\text{atmosphere}|r) \cdot q_{\text{atmosphere}\rightarrow\text{storm}}] \\ &\quad + q_{\text{atmosphere}\rightarrow\text{storm}} \cdot [P(\text{atmosphere}|r) - P(\text{atmosphere}|\bar{r})].\end{aligned}$$

Thus, with $q_{\text{reading} \rightarrow \text{storm}} = 0$, as it should, ΔP_r would nonetheless have a high value, because of a high value in the last product ($q_{\text{atmosphere} \rightarrow \text{storm}}$ has a high value, and r covaries with a drop in atmospheric pressure).

Another difference between covariation and causation explained by the incorporation of generativity in the power PC theory concerns an asymmetry in causal inference for the same covariation due to the directionality of causality (Waldmann & Holyoak, 1992): Causes compete, but effects do not. It has been shown that causal ratings of candidate c with respect to effect e are reduced when an alternative cause that covaries with c also covaries with e, but the diagnostic ratings of c with respect to e remain unchanged when an alternative effect that covaries with e also covaries with c. The asymmetry is due to the assumption of directionality inherent in generativity. A cause generates its effect, not vice versa; the causal arrow points in one direction. Our derivations (Cheng, 1997; Novick & Cheng, 2004) show that alternative causes, but not alternative effects, need to be controlled. The reduction in causal ratings in the causal condition in the study by Waldmann and Holyoak (1992) results from controlling for alternative causes, which has no analogue in the diagnostic condition.

An analysis of Shultz's (1982) findings. Within our framework, in which covariation is to generativity as evidence is to

[4] Following Cheng (2000), we use q_c to represent the generative causal power of candidate c.

theory, a contest between generativity and covariation is nonsensical. Theories do not defeat evidence used for evaluating them. Why, then, would White (2005) conclude from Shultz's (1982) findings that generativity wins over regularity? Contrary to claims by Shultz and White that the generativity inferred by participants in these experiments could not be regularity based, in fact it was, without exception. As we explain, Shultz's results support rather than contradict the power PC theory.

White (2005) offers Shultz's Experiment 4, conducted on children from a rural area of Mali, as convincing evidence that people form causal judgments on the basis of generativity rather than covariation. The experiment had two parts: learning that vibrating tuning forks can cause a column of air to resonate and learning that lamps can make a bright spot on the wall unless their light is blocked.

Notably, for the tuning forks task, for which the participants were unlikely to have relevant prior causal knowledge, the experiment had a training phase. During this phase, Shultz (1982) taught his participants the novel generativity by presenting none other than covariation information! He wrote,

> Because few children were familiar with the principles of sound transmission, it was considered necessary to provide a brief training phase. . . . The child was asked to feel the prongs of each fork before and after the experimenter banged it [no vibration in fork is paired with no banging; vibration is paired with banging]. . . . The experimenter . . . also demonstrated that each vibrating fork could make the box ring when placed in front of the open end [vibrating fork in front of opening is paired with ringing, absence of vibrating fork in front of opening is paired with no ringing]. Presumably, this training phase conveyed something of the vibratory nature of sound. (Shultz, 1982, pp. 8–10)

(Notice that the vibratory nature of sound, being a conclusion rather than the input, does not contradict regularity accounts.) Contrary to Shultz's intention, what he presented to the participants was clearly covariation information.[5] The inclusion of the training phase shows that Shultz implicitly recognized that there was no choice but to use covariation, even though his explicit goal was to refute the need for it.

For the lamps and light-on-the-wall task, the children were likely to have relevant prior causal knowledge involving other light sources. They would have seen light from a bonfire, candles, or the sun that would covary with surfaces being lit up (i.e., they would have causal knowledge of light inferred from regularity information). Notably, in judging which lamp lit the spot on the wall, the children were asked to choose between (a) a lamp that was observed to be on when turned around right in front of them, without interruption (thus the lamp was likely to have been on when it faced the wall), and (b) a lamp that had the same paint color (red) as one that was on during a previous stage, with the stages separated by the introduction of a screen that blocked the children's view of the lamps (thus there is less assurance that this lamp is on). Participants might realize that in between the stages, out of their sight behind the screen, there were plenty of opportunities for alternative causes to act. In our terms, this was a contest between (a) generativity based on unconfounded covariation based on information obtained prior to the experiment—that between the lit light facing the wall and the bright spot on the wall—and (b) generativity based on potentially confounded covariation presented during the experiment—that between the red color of a lamp and the bright spot on the wall. Option b is more likely to be confounded and therefore noncausal. Participants' preference for the first option supports our theory, which predicts the use of the criterion of "no confounding" for differentiating causation from mere covariation.

Causes Versus Enablers

Cheng and Novick's (1991) explanation. White (2005) claims that our approach cannot explain the distinction between causes and enabling conditions. Our theory in fact does offer such an explanation. We quote from Cheng's (1997) summary of our probabilistic contrast model: Under our approach,

> candidate *i* is an enabling condition for a cause *j* if *i* is constantly present in a reasoner's current focal set [the primary set of events used by the reasoner to make the causal judgment in the current context] but covaries with the effect *e* in another focal set, and *j* no longer covaries with *e* in a focal set in which *i* is constantly absent. . . . To illustrate . . . because fire occurs more frequently given the dropping of a lit cigarette than otherwise, the dropped cigarette is a cause. Oxygen, however, is present in all forests. Its contrast therefore cannot be computed within the current focal set. It is not causally irrelevant, however, because it does covary with fire in another focal set, one that includes events in which oxygen is absent as well as those in which it is present (e.g., in chemistry laboratories). Oxygen is therefore an enabling condition. Finally, it is an enabling condition rather than an alternative cause because, in yet another focal set in which oxygen is always absent (e.g., also in chemistry laboratories), a lit cigarette no longer covaries with a bigger fire. (p. 372)

The above explanation remains intact for the power PC theory, except that the covariation evaluations would be replaced by causal power evaluations.

Cheng and Novick (1991) tested these predictions. Consider their Experiment 2, which examined reasoners' causal analyses of two scenarios concerning plant growth. In one, sunlight covaried with growth, and nutrients and water were constantly present. In the other, nutrients covaried with growth, and sunlight and water were constantly present. In strong support of our predictions, 91% of participants identified the covarying factor as the cause (sunlight for one scenario, nutrients for the other), and 83% identified the constant factors as enablers (nutrients for one scenario, sunlight for the other, water for both). Thus, as proposed by the power PC theory, the identification of candidates as causes versus enablers depends on covariation within focal sets.

---

[5] The training phase was repeated with slight variations, all presenting participants with covariation information. For example, in one of the subsequent variations, the experimenter, as before, "banged each fork in turn and placed it in front of the opening to make the box resonate. While it was resonating, she blocked the sound wave by inserting a sheet of cardboard between the box and the fork, thus stopping the resonation" (Shultz, 1982, p. 8). Thus, a vibrating fork in front of the opening without any sheet of cardboard in between is paired with ringing, and a vibrating fork in front of the opening with an intervening cardboard is paired with no ringing. Note that the causal interpretations in Shultz's (1982) description, such as, "to make the box resonate," "she blocked the sound wave," and "stopping the resonation," were not what the participants observed as input and would be fine as output.

White's (2005) alternative explanation. White (2005) reviews his experiments (White, 2000), testing his alternative hypothesis that a common factor (CF; one that is present in all instances in a set) should be judged as the cause of an effect in question, whereas a *covariate*, a factor that covaries with the effect, is merely an enabling condition. He interprets his results as supporting his hypothesis. If his interpretation is accurate, all previous regularity accounts, including the power PC theory, would be refuted. But, there is a major theoretical problem with White's hypothesis; it represents an explanatory move backward from even the contingency model (Jenkins & Ward, 1965). According to his hypothesis, a car running a red light and hitting a pedestrian is not the cause of the pedestrian's injury; because it is a covariate, it is merely an enabler.[6] Furthermore, when would there be preventive causes? They would be impossible; disablers would presumably covary negatively with the effect, but preventive causes cannot be factors that are absent on every trial.

In view of the implausibility of White's (2000) results, we carefully examined his experiments to see if there might be an alternative explanation. Our analysis reveals that his data do not allow discrimination between the contradictory predictions of his approach and of the power PC theory. We briefly review two problems with the experimental methodology.

First, throughout his experiments, potentially critical aspects of the experimental materials were not counterbalanced. The effect in question was an allergic reaction, and the potential causes were food additives. For Experiments 1 and 2, the CF was always "sodium trisulphate," and the covariate was always "Wassmeier's salts"; it is possible that the uncounterbalanced names were to different degrees suggestive of the additives being allergenic. For example, sodium trisulphate, the CF that White (2000) predicts to be a cause, may be more suggestive of toxicity than Wassmeier's salts, the covariate that he predicts to be the enabler. (College students presumably know that the human body naturally contains various salts.) For Experiments 3 and 4, the names of the food additives corresponding to the CF and the covariate were not reported; in any case, they were not counterbalanced. Another extraneous factor that might have contributed to the results is that throughout the materials, the CF food additive was always listed first, before the covariate, which might have subtly suggested that this factor was more important (i.e., more causal). These failures in counterbalancing alone may explain the critical aspects of White's results.

Second, White (2000) assumed that the participant's focal set for inference necessarily is identical to the set of instances presented by the experimenter. This assumption seems dubious. Participants were likely to have used the information, from their prior knowledge, that on occasions during which there is no consumption of food, the probability of an allergic reaction is generally low. If so, the CF, being absent on these occasions, would not be a constant factor in the participants' focal set.[7] Novick, Fratianne, and Cheng (1992) found that participants "fill in," on the basis of their prior knowledge, relevant information that is missing from the explicit experimental context and use the combined set of regularity information to guide their causal inferences.

Several problematic consequences follow from the ambiguity in defining the focal set for causal inference. First, the supposed CF cannot be assumed to be constant in the participants' focal set; in fact, it is likely to be a covariate, contrary to White's (2000) intended design. Second, because there is no constant factor, the problems do not include what would be identified as an enabling condition by the power PC theory. One cannot evaluate whether reasoners determine causal status according to one theory or another unless the relevant variables for both theories are instantiated in the experimental materials. Likewise, for many of White's problems with no CF, including the one that White (2005) identifies as critical for discriminating between the two theories, the response predicted by the power PC theory is that the two additives interact; this response option was not among those offered to participants in any of the four experiments.

The Influence of Prevalence on Causal Judgments

White's (2005) third line of evidence against regularity accounts concerns the influence of prevalence on causal judgments (Johnson, Boyd, & Magnani, 1994; Johnson, Long, & Robinson, 2001; White, 2004). He argues that Johnson and colleagues' (1994, 2001) finding that compared with a rare factor, a more prevalent factor that covaried less with an effect received a higher causal judgment "appears to contradict the predictions of any regularity-based model, including the power PC theory, because no such model would predict that a weaker covariate would be given a higher causal judgment than a stronger covariate" (White, 2005, p. 681). He also interprets the influence of prevalence per se as a refutation of the regularity accounts. Johnson et al. draw a similar conclusion.

We show that these interpretations are incorrect: Because Johnson et al.'s (1994, 2001) studies asked participants questions on *causal attribution* (e.g., what proportion of the occurrence of e is due to c?) rather than on inferred causal strength (e.g., how often does e occur when c occurs, in the absence of other causes of e?), the reported influences of prevalence are in fact predicted by our power PC approach. The answers to causal-attribution and causal-strength questions, by the very nature of those questions, should be different even for the same causal relation. Moreover, conflating prevalence, $P(c)$, and causal strength, q_c, loses useful information. One reason for distinguishing between them is that $P(c)$ is observable. In transfer situations in which $P(c)$ is different from that observed in the learning context, if one represents q_c separately from $P(c)$, one can still predict the probability of the effect explained by c, by instantiating the new value of $P(c)$ in the product, $q_c \cdot P(c)$. We discuss the role of prevalence in causal attribution in more depth in the next section.

---

[6] The lights in the heels of the pedestrian's shoes, on the other hand, which flashed each time the pedestrian took a step, including the step the pedestrian took as the car hit him or her, qualifies as the cause because it is a CF in the situation as described.

[7] Moreover, consider White's (2000) experimental situation: "Imagine that you are a doctor investigating patients suffering from severe allergic reactions. You are trying to find out whether their allergic reactions are caused by substances in the food they eat.. . . You ask your patients to eat a certain number of meals in which the food additives are either present or absent" (White, 2000, p. 1087). Participants, taking on the persona of the doctor as requested, might have assumed that the doctor used his or her knowledge of the patient's prior history to select which additives to investigate and to assign them to particular meals. The more frequently assigned food additives might be assumed to be the more likely suspects.

In this section, we consider Johnson et al.'s (1994, 2001) results in more detail. Our interpretation is that Johnson et al. created a complex scenario involving a causal chain with an intermediate node that has a threshold. When participants were asked to apportion the occurrence of an outcome to the target causes, those at the beginning of the causal chain, they did so according to the contribution of each target cause to the magnitude of the intermediate node because that node is what in turn caused the outcome. Participants were informed that the prevalent factor contributed more to the intermediate node than did the rare one. It follows that they would attribute the outcome to the prevalent factor more than to the rare one. By the nature of the threshold, a prevalent factor would covary less with the node reaching its threshold than would a rare factor; but participants were not asked to estimate causal strength (i.e., how often a factor causes the intermediate node to reach threshold when the factor is present).

To explain why Johnson et al.'s (1994, 2001) scenarios involve a threshold, we need to explain why their scenarios do not involve a conjunctive cause. Consider the following scenario from their experiments, in which the prevalence of two factors varies:

> Jean is the kind of person who is bothered by feelings of workload-related stress most of the time. She is bothered by feelings of workload-related stress considerably more often than she is bothered by feelings of annoyance with her coworkers. (Johnson et al., 2001, p. 402)

Jean makes mistakes at work only when she experiences both workload-related stress and feelings of annoyance with coworkers. If two factors (workload-related stress and annoyance with coworkers) are truly jointly necessary for the effect (making mistakes) and interact to produce it, then it simply would make no sense to ask how much each factor contributes to the effect—they are both necessary! A threshold conception of the situation, however, would render Johnson et al.'s questions sensible for the data presented: The intermediate node, "feeling bothered," has a threshold below which "making mistakes" (the effect) does not occur and above which it can occur. (White, 2004, made a similar suggestion.) Under this conception, the target factors are actually additive; they do not interact as do factors in a conjunctive cause. The additivity in this model makes the division of contribution meaningful—the factor that is responsible for more "units" along the continuum of "feeling bothered" would contribute more.

Causal Attribution Versus Causal Strength

By causal attribution, we mean the apportioning of the observed probability of an effect e, $P(e)$, according to causes that contributed to its occurrence. Notice that by the nature of this question, $P(e)$ is the denominator. (When asking how much of the occurrence of e is due to c, one has to be asking about e when it occurred.) In contrast, by causal strength (or causal power), we mean the probability that a candidate cause c produces e when c occurs (e.g., Cartwright, 1989; Cheng, 1997). By the nature of this question, $P(c)$ is in the denominator.[8]

Both Johnson et al. (1994, 2001) and White (2004) asked participants a causal attribution question. In Johnson et al. (2001, p. 405), participants were asked how much each candidate cause (e.g., workload stress and coworkers' behavior) contributed to the intermediate cause of mistakes (feeling bothered)—that is, which candidate was "the more important cause" of the behavior. White's (2004) experiments asked to what extent each food additive caused the patients' headaches. That is, regarding patients who have headaches, to what extent are their headaches due to additive X? For such questions, prevalence can be a normative influence in addition to strength, as we show in the following computational analysis.

We consider three alternative measures of causal attribution involving causes and effects that are binary variables.[9] The measures differ because attribution questions involving different givens (i.e., input and assumptions) ought to have different answers. The answers are all derived under our power PC framework: In this particular case, they assume either that alternative causes have no interaction with c (Cheng, 1997) or that the interacting factors occur with the same probability across the learning and transfer contexts (Cheng, 2000).

We begin with some notation and relevant definitions. Let $c \rightarrow e$ denote that e is produced by c. We use $P(c \rightarrow e|e)$ to represent the probability that e is due to c given that e has occurred and q_c to represent the causal power of c (Cheng, 1997; Novick & Cheng, 2004). Because c and a, the composite of all causes alternative to c, are the only causes of e, e is (nonexclusively) produced by c or by a. Given the relevant independence assumptions, it follows that (Equation 2 in Cheng, 1997)

$$P(e) = P(c) \cdot q_c + P(a) \cdot q_a - P(c) \cdot q_c \cdot P(a) \cdot q_a. \qquad (1)$$

The first two terms on the right-hand side are, respectively, the probability that e occurs and is produced by c and the probability that e occurs and is produced by a. The remaining term is the probability that e occurs and is produced both by c and by a. It also follows that $P(e|\bar{c})$ estimates $P(a) \cdot q_a$; this is because c occurs independently of a and, in the absence of c, only a produces e. Therefore,

$$P(e) = P(c) \cdot q_c + P(e|\bar{c}) - P(c) \cdot q_c \cdot P(e|\bar{c}). \qquad (2)$$

Now, here are the three causal attribution measures:

1. Attribution to c: In this situation, one knows that c has occurred with probability $P(c)$. For cases in which e has occurred, one might ask, how often is e due to c? As just explained, $P(e)$ is in the denominator. Because e is caused by c with probability $P(c) \cdot q_c$, the answer is

$$P(c \rightarrow e|e) = \frac{P(c) \cdot q_c}{P(e)}. \qquad (3)$$

Note that the output of this function is a probability—namely, $P(c \rightarrow e|e)$—the proportion of the occurrences of e that are caused by c.

2. Attribution to c alone: In this situation, everything is the same as before, except that the question is how often is e due to c alone

[8] Consistent with what causal strength means in our theory (as opposed to how to estimate this quantity from observations), q_c is equal to how often e is produced by c divided by how often c occurs: $\frac{P(c) \cdot q_c}{P(c)} = q_c$. Causal strength and causal attribution have the same "theoretical" numerator: $P(c) \cdot q_c$.

[9] We thank Clark Glymour for discussion on these measures.

(i.e., not due to other known and unknown causes of e). From Equation 2, one sees that one must subtract $P(c) \cdot q_c \cdot P(e|\bar{c})$, the probability that e is produced both by c and by a, from $P(c) \cdot q_c$ to obtain $P(c\text{-alone} \rightarrow e)$. The answer in this case would be

$$P(c\text{-alone} \rightarrow e|e) = \frac{P(c) \cdot q_c - P(c) \cdot q_c \cdot P(e|\bar{c})}{P(e)}$$

$$= \frac{P(c) \cdot q_c \cdot [1 - P(e|\bar{c})]}{P(e)} = \frac{P(c) \cdot \Delta P}{P(e)}. \quad (4)$$

The simplification in the last step makes use of the equation specifying simple causal power (Cheng, 1997), contextual power (Cheng, 2000), or probability of sufficiency (PS; Pearl, 2000):

$$q_c = \frac{\Delta P}{1 - P(e|\bar{c})} = \frac{P(e|c) - P(e|\bar{c})}{1 - P(e|\bar{c})}. \quad (5)$$

3. Attribution to c, given that c and e both occurred: Finally, consider how often e is due to c, but this time, not only does one know that e has occurred, one also knows that c has occurred.

From the definition of $P(e|c)$,

$$P(c,e) = P(c) \cdot P(e|c). \quad (6)$$

As before, e is due to c with probability $P(c) \cdot q_c$. Therefore,

$$P(c \rightarrow e|c,e) = \frac{P(c) \cdot q_c}{P(c,e)} = \frac{P(c) \cdot q_c}{P(c) \cdot P(e|c)} = \frac{q_c}{P(e|c)}. \quad (7)$$

Notice that $P(c)$, the prevalence of c, occurs in the first two measures but not the third. White's (2004) and Johnson et al.'s (1994, 2001) questions concerned the first measure: Participants were asked about attribution to c in general rather than about attribution to c alone or about cases in which c has occurred. As Equation 3 shows, there should be an influence of prevalence for their attribution question; that is, the influence of prevalence is in fact normative.

Conclusions Regarding White's (2005) Comment

Without a computational account, there would be no framework within which to judge whether a finding contradicts a theory (e.g., whether an influence of prevalence contradicts covariational theories). Neither would there be nonarbitrary constraints on the construction of experiments and their materials (e.g., see our discussion of Shultz, 1982). Moreover, there would be no glue to hold together explanations of various types of causal judgments; for example, in White's (2005) theory, why is a purely covariational model, consisting of constant factors and covariates, part of an antiregularity causal powers theory? Within our power PC framework, however, extensions to causal roles and to various causal attribution questions are natural and logically consistent. The framework offers a coherent explanation of people's flexible causal judgments across disparate situations.

An Analysis of Luhmann and Ahn (2005)

Overview

Luhmann and Ahn (2005) argue that when reasoners make the power PC assumptions (listed on p. 686 in their article for simple, generative causal power), if they further assume causal determinism, then the causal power of the candidate cause will be either 0 or 1, and there would be no need to use the equations in our theory. Luhmann and Ahn advocate a deterministic view that explains probabilistic causal powers by incomplete knowledge. In addition, they criticize the assumptions underlying our theory for being unrealistic and lacking individual validation, and they note an apparent contradiction between estimating simple causal power (Cheng, 1997) and estimating contextual causal power (Cheng, 2000). Finally, they exclude causal power as an interpretation of human causal learning and deny generalizability beyond the learning context as a goal of causal inference. Instead, they present Pearl's (2000) PS and the "intentional" estimation of contextual power as potential solutions to the supposed problems confronting our theory.

We explain that Luhmann and Ahn's (2005) conclusion that causal power is either 0 or 1 rests on their hidden assumption that the reasoner's hypothesized cause is always perfectly correct. We show that their argument that an imperfect hypothesized cause results in confounding, and hence a violation of the no confounding assumption, leads to a logical contradiction and the paralysis of causal learning. We further explain that their other criticisms of our theory are due to (a) their failure to take into account a key goal of causal inference—predicting the consequences of actions—and (b) their implicit assumption that the boundary of the learning context is always or typically identifiable. Finally, we note some untenable demands they make of a theory of causal reasoning.

Probabilistic Causality Due to a Reasoner's Suboptimal Representation of the Cause

The reasoner, not being omniscient, may entertain a partially correct candidate cause. We show that a probabilistic causal power can be obtained when all of the power PC assumptions are met if candidate cause c is an imperfect hypothesis, even for a reasoner who assumes causal determinism.

A hypothetical example. Let us consider a simple case in which a reasoner has a suboptimal representation of the cause. We assume for the sake of argument, along with Luhmann and Ahn (2005), that the reasoner believes in causal determinism. (We note, however, that contrary to what they seem to suggest, not all reasoners hold this belief; in fact, half of the authors of this article believe that there is inherent randomness in the world.) Suppose, for the purpose of this illustration, that some conjunction of substances in dried orange peel deterministically repels beetles, and no other fruit peel (either alone or in combination with another factor) has this effect. The reasoner's hypothesis, however, is that dried peels from citrus fruits repel beetles; thus, the reasoner's candidate cause is an overly general category with respect to the true cause. To test this hypothesis, the reasoner cuts up the peels of three oranges, three lemons, and three apples (each fruit being roughly equal in size) and places the pieces of peel from each fruit in a separate compartment of a drying rack. After the peels have dried out completely (at which point the pieces from the two types of citrus fruit are indistinguishable), the reasoner puts the peels of each fruit in a separate bag, labels each bag as "citrus" or "apple," and tests each bag on beetles.

Further suppose that the reasoner (correctly) makes all the causal power assumptions: In particular, there are no preventers of the effect for the population of beetles under study, and citrus peels do not interact with some other factor to repel beetles. Under these assumptions, the causal power for the candidate would be greater than 0 (three of the six bags of citrus peels would repel beetles) but less than 1 (not all bags of citrus peels would repel beetles); specifically, $q_c = 0.5$.[10] Luhmann and Ahn's (2005) central thesis is thereby refuted: Even when all the causal power assumptions are met, allowing for the possibility of an imperfect hypothesis will explain apparent probabilistic causal powers for a reasoner who believes in determinism. Thus, contrary to Luhmann and Ahn's assertion, arriving at a causal power between 0 and 1 need not indicate any violation of the power PC assumptions. We have illustrated one way of having an imperfect hypothesis; there are other ways (e.g., the true cause in our example is instead both dried and undried orange peels, so that the reasoner's candidate cause now partially overlaps with the true cause but does not include it as a subset; see Klayman & Ha, 1987; Lien & Cheng, 2000).

Alternative hypotheses versus alternative causes. Luhmann and Ahn (2005) would dismiss our refutation of their thesis, however. According to them, people should withhold causal judgment when a candidate cause c is represented overly generally (i.e., the true cause is a subset of c), because c would be "confounded" with the true cause (a subset of c "occurs" more often when c occurs than when c does not occur). Notice that Pearl's (2000) PS and Cheng's (2000) contextual power also require no confounding; the prediction to withhold judgment therefore should likewise apply. Luhmann and Ahn's objection confuses *alternative hypotheses* regarding a candidate cause (e.g., dried citrus peels vs. dried orange peels) with *alternative causes* of e that jointly explain the occurrence of e (e.g., say that dried orange peels, catnip, garlic, etc., repel beetles in our hypothetical situation). Treating a subset of the candidate cause as an alternative cause is not a coherent interpretation of our theory. Below, we illustrate the incoherence using Equation 1 in Luhmann and Ahn (from Cheng, 1997), an equation that logically follows from the power PC assumptions.

In our theory, all direct causes of e are "partitioned" into the candidate(s) on one hand and the composite of alternative causes on the other (e.g., see Novick & Cheng, 2004, p. 459 for simple power and p. 462 for conjunctive power). The two sets are therefore mutually exclusive. For example, for the outcome "lung cancer," if the candidate cause being evaluated is "inhaling tobacco smoke," the composite of alternative causes includes everything else (i.e., other than tobacco smoke) that causes lung cancer (exposure to asbestos, working in a coal mine, etc.). Together, the candidate and the alternative composite jointly explain the occurrence of e. They contribute *simultaneously* to the instantiation of a causal power equation. Using our beetle repellent example to be concrete, for Equation 1 in Luhmann and Ahn (2005), i is "dried citrus peels" (including dried orange peels), and all other beetle repellents (catnip, garlic, etc.), known or unknown, are part of a (for the nine bags in our example, the reasoner's focal set in this situation, these alternative causes happen to be absent, so that $P(a) = 0$). In contrast, alternative hypotheses regarding a candidate cause map onto *different* instantiations of a causal power equation. For example, in a different instantiation of Equation 1, i might be "dried orange peels," and dried lemon peels would no longer belong in the category.

Let us now instantiate Equation 1 in Luhmann and Ahn (2005) using their "alternative hypotheses" interpretation of alternative causes. For our beetle repellent example, if i is "dried citrus peels" and the true cause is "dried orange peels" (as in our original example), then according to Luhmann and Ahn's treatment of the true cause as an alternative cause (i.e., as a in the equation, assuming that the true cause is the reasoner's only alternative hypothesis in the study involving the nine bags), the left-hand side of Equation 1, $P(e|i)$, would be 0.5 (the three orange bags of the six citrus bags repel beetles). But, the right-hand side, $q_i + P(a|i) \cdot q_a - q_i \cdot P(a|i) \cdot q_a$, would be $0.5 + 0.5 \cdot 1 - 0.5 \cdot 0.5 \cdot 1 = 0.75$, contradicting the left-hand side.[11] This contradiction reveals the incoherence of Luhmann and Ahn's "confounding-by-subset" argument. Critical concepts such as "alternative causes" and "alternative hypotheses" may seem verbally confusable, but they have obvious, distinct, and logically consistent operational definitions in our theory, as we have shown.

Notice that confounding by subsets, unlike the kind of confounding that matters in causal inference, concerns category membership rather than occurrence. Apple peels cannot be orange peels (they are mutually exclusive in that sense), but they can certainly co-occur with orange peels in an agent (e.g., in a potpourri); orange peels are citrus peels (one is a subset of the other), and it would be anomalous to say that orange peels co-occur with citrus peels.

Paralysis of causal inference. Luhmann and Ahn's (2005) confounding-by-subset argument also suffers from more general problems. First, how would the reasoner, who obviously does not already know the true cause, tell when there is "confounding" with a true cause? Luhmann and Ahn seem to confuse the desired output of causal learning with the input. Moreover, all causes are "confounded" (i.e., correlated in category membership) with their subsets or supersets, and the reasoner, not knowing the true cause, would never be able to rule out the possibility that a subset or superset is the unknown true cause. Thus, no causal inference should ever occur.

Even from the perspective of an omniscient being who is judging human inference, causal inference would be justified only if the reasoner hypothesizes a candidate that is 100% accurate, so that the candidate is the true cause (assuming, reasonably, that a perfect correlation between hypotheses, unlike that between causes, does not count as confounding). All other candidates would risk a correlation between the candidate and the true cause, and hence there would be "confounding." There is no analogue of controlling for alternative causes: One cannot "manipulate" a certain hypothesis and thereby make it independent of the true cause. Even when the candidate is 100% wrong, judgment should be withheld because the candidate would be negatively correlated

[10] Instantiating our Equation 5 with the observed probabilities in this situation, we get $q_c = \dfrac{P(e|c) - P(e|\bar{c})}{1 - P(e|\bar{c})} = \dfrac{0.5 - 0}{1 - 0} = 0.5$.

[11] Notice that confounding is irrelevant with respect to these calculations because no confounding is not a requirement for Equation 1. Also notice that in this instantiation, we limited the number of alternative hypotheses to one. But, the number of hypotheses a reasoner entertains is limited only by his or her imagination; it is therefore in general unclear what would be meaningful estimates of $P(a|i)$ under Luhmann and Ahn's (2005) interpretation.

with the true cause.[12] Thus, from both perspectives, knowledge of the true cause would be required for causal learning.[13] Given that the reasoner does not know the true cause, the only world then in which causal inference could occur would be one in which events have no causes (so that confounding with the true cause can be ruled out), a completely probabilistic and noncausal world.

But, the possibility that the candidate cause does not map perfectly onto the true cause does not and should not stop causal inference. Suppose, for example, that the reasoner successfully manipulates whether candidate *c*, inhalation of tobacco smoke, occurs for some poor laboratory rats to test whether it causes lung cancer. Regardless of the relationship between *c* and the true cause (say, inhaling smoke from only tobacco plants with gene *x*), the reasoner is likely to, and should, interpret the covariation between *c* and *e* as causation. That is what the experimental method allows. Clearly, knowledge of the true cause is not part of the input, or even the output, of causal inference. Our power PC theory, which makes a distinction between category membership (belonging to *i*, the candidate-cause category, vs. *a*, the alternative-causes category) and occurrence ($P(i)$ and $P(a)$), avoids the paralysis implied by Luhmann and Ahn's (2005) conception of "alternative causes."

At first glance, it may appear that some of the confounding examples discussed earlier, such as White's (2005) pill example, fit Luhmann and Ahn's (2005) confounding-by-subset argument. In fact, they do not. The pill, for example, is not a superordinate category of the medicine in the pill (e.g., the medicine is in every pill; it can be packaged in a liquid form for injection). Instead, the pill is the agent, an agent that may contain other causes of the outcome (e.g., the cause of the placebo effect). If there happens to be no alternative causes in the pill, causal inference regarding the medicine would be fully justified. Confounding is due to other causes covarying with the candidate with respect to the patients as a result of the distribution of the agents (e.g., medicine and the cause of the placebo effect co-occurring in people who took the pill and neither occurring in those who did not take the pill; see the *Causal Learning as Problem Solving: Causal Roles as Outputs* section for definitions of patient and agent) rather than due to a subset of members of the candidate-cause category covarying with the full category. Confounding by alternative causes and overlaps in category membership are of course not mutually exclusive.

In summary, Luhmann and Ahn (2005) confuse (a) alternative causes with alternative hypotheses, leading to a logical contradiction, and (b) the desired output of causal learning with the input, resulting in an unattainable and needless demand on the reasoner. Contrary to their assertion, a probabilistic causal power need not indicate any violation of the power PC assumptions, even for a reasoner who believes in causal determinism (their discussion of causal determinism is therefore irrelevant). A probabilistic power might instead reflect the reasoner's imperfect representation of the cause. Depending on the reasoner's purposes and resources, this might spur him or her to seek a more accurate representation.

Causal Power as Ideal Under the Broader Goals of Explanation and Prediction

Luhmann and Ahn (2005) write,

> the power PC theory lays out several assumptions, each of which is a necessary condition for computing causal power from observations. Thus, if any one of them is violated, one should withhold his or her judgment. . . . In contrast, the contextual power theory appears to state that when assumptions are violated, people would still compute something useful, namely, contextual power. (p. 680)

The prediction to withhold judgment (as opposed to drawing a definite conclusion) in view of a violation of an assumption also forms the basis of Luhmann and Ahn's proposed empirical test between Pearl's (2000) PS and simple causal power. We argue that Luhmann and Ahn's views on these issues might be explained by their apparent failure to (a) consider the prediction of the consequences of actions as a goal of causal inference and (b) recognize that contexts often do not carry labels.

On the compatibility of context-dependent and context-independent causal inference. It seems quite odd to us for a sharp conceptual distinction to be drawn between Cheng's (2000) contextual causal power and our context-independent measures (Cheng, 1997; Novick & Cheng, 2004). To us, aiming for causal power and accepting contextual power is as "contradictory" as aiming for a gold medal and accepting silver. Luhmann and Ahn's (2005) argument omits Cheng's (2000) consideration of the other conditions under which estimates of causal power would be useful for predicting the consequences of actions, a key goal of causal learning for the obvious reason that it supports flexibly adaptive goal-directed actions (see the first paragraph in Cheng, 2000).[14] They narrowly interpret inferring context-independent causal relations as a goal in itself, divorced from the broader goal of predicting the consequences of interventions. Our view instead is that it is an ideal within that broader goal.

To draw a conclusion from a simple causal power equation means arriving at a definite value (or a restricted range of values) for the expression on the right-hand side of Equation 5. The power PC assumptions listed in Luhmann and Ahn (2005) are necessary elements of a set that is sufficient for deriving these expressions. But, Cheng (2000) showed that there are other sufficient sets for deriving the same expressions (e.g., replacing the independent-influence assumption with the assumption that the background factors that interact with the candidate to produce the effect occur with the same probability in a new context), implying that none of the sets is necessary for making predictions on the basis of the expressions. Thus, our approach does not imply that judgment should be withheld if any of the simple power assumptions is

[12] Suppose the hypothesized cause is now completely wrong, apple peels for the beetle example, with orange peels being the true cause as before. The hypothesized cause is negatively correlated with the true cause because the probability that the peels are orange given that the peels are apple is 0, but the probability that the peels are orange given that the peels are not apple is positive (0.5 in our example).

[13] Luhmann and Ahn's (2005) discussion of determinism for conjunctive causal power (Novick & Cheng, 2004) likewise begins with the assumption of a perfect hypothesis.

[14] Luhmann and Ahn (2005) quote Cheng (2000) on this goal but ignore it. They also quote Cheng's (2000) statement that when the no-preventive-background-cause assumption is violated, "there is no unique solution [for causal power] in general" (p. 239). This quote may be misleading when taken out of context. Cheng (2000) wrote in the immediately following sentence, "it turns out that [the power PC equation] nonetheless provides a conservative, and hence useful, estimate for predicting [the consequences of] interventions with *i*" (middle of p. 239).

violated. It is withheld only if no causal judgment is possible under any of the models a reasoner is willing to entertain (e.g., see Wu & Cheng, 1999). Aiming for context-free causal relations merely means a preference for a fuller understanding. It is puzzling that Luhmann and Ahn (2005) object, especially so vociferously, to having a goal of inferring context-free causal relations even as they advocate determinism. Only when causal relations are fully and accurately specified could they be deterministic. Thus, in their scheme, one is forbidden to aim for the only kind of explanation in which one believes.

Luhmann and Ahn (2005) propose a possible solution for the apparent contradiction between contextual causal power and the goal of inferring context-free causal relations: Reasoners estimate contextual causal power without awareness. Their argument is analogous to suggesting that to avoid a contradiction with aiming for a gold medal, one would need to accept a silver medal without awareness that it is silver. In fact, estimating Cheng's (2000) contextual causal power, intentionally or not, is not only consistent with our view, it is an integral part of the power PC approach. We view contextual power and causal power, as well as Pearl's (2000) PS and the various causal attribution measures discussed earlier in our reply to White (2005), as compatible measures whose applicability depends on the causal question, the situation, and the reasoner's beliefs and utilities (e.g., the cost of making a prediction error).

On the need to generalize across contexts. For everyday causal learning, drawing a sharp conceptual distinction between context-dependent and context-independent inference, and between whether the use of contextual causal power is intentional or unintentional, is both odd and immaterial, because contexts in everyday life seldom carry a label. By *context*, we mean the states of background causes of a target effect *e*, that is, causes of *e* other than the candidate(s) (Cheng, 2000, p. 227). Although new contexts are sometimes marked by new or different information (e.g., an interacting factor becomes known and measurable; the base rate of *e* changes), old contexts are never marked: It is inherently impossible to check for potential changes in the probability of *unobserved* interacting or preventive factors. Scientific studies circumvent that problem by random sampling. The sampling procedure defines the boundaries of the learning context: the population from which the random sample is drawn. But, everyday events are seldom randomly sampled. The unidirectionality of time, for one thing, makes random sampling of events difficult. Given that no new experience is ever identical to the learning one, forgoing generalization to new contexts, as Luhmann and Ahn (2005) advocate, would imply no predictions regarding future interventions. The perfectionism would lead to paralysis. Given that one generally has practical purposes to achieve, however, there is often no choice but to risk generalization to a potentially new context and then to try to modify one's causal model if it is proven wrong. When one travels to a new country, for example, should one not assume that washing with soap kills germs, even though it is possible that some interacting factor in the water occurs with a different probability?

The nature of the assumptions in our theory. In our view, under the broader goals of explanation and prediction, and given the typically unmarked boundaries of the learning context, the reasoner begins with simple assumptions as tentative "working hypotheses" (Novick & Cheng, 2004, p. 471); these assumptions reflect what the reasoner is willing to believe in a situation. Even as reasoners draw inferences, however, they allow for the possibility that their assumptions are wrong. If any assumption is believed to be unrealistic, more complex and realistic assumptions will be made, allowing reasoners to "incrementally construct a picture of causal relations in a complex world" (Cheng, 2000, p. 227).

It should be clear, therefore, that our theory does not propose a fixed set of assumptions. Nevertheless, Luhmann and Ahn's (2005) characterization of our view and their criticism of the unrealistic nature of our assumptions in particular may suggest that a certain fixed set of assumptions is what sets our theory apart from previous covariational models. Contrary to their characterization, what is new about our theory, pervading the 22 sets of assumptions we have considered, is the explanation of observations by unobservable causal powers (i.e., the incorporation of domain-free generativity; see Cheng, 1997, 2000; Novick & Cheng, 2004). This added explanatory layer in our theory is important in multiple ways: (a) It is what allows the differentiation between covariation and causation; (b) it allows a logically consistent explanation of a variety of causal judgments (as we illustrated earlier in our reply to White, 2005); (c) it provides a nonarbitrary filter for illogical arguments (as we illustrated with Luhmann & Ahn's, 2005, deterministic causal power argument); (d) it is what allows Novick and Cheng (2004) to show that the cross-product ratio commonly used as a criterion of independence is in fact arbitrary; and (e) it is what makes a Bayesian account of structure learning (Griffiths & Tenenbaum, in press; Tenenbaum & Griffiths, 2001) causal. The sets of assumptions in Cheng (1997) and Novick and Cheng (2004) are special only in that they are the simplest possible for their respective purposes.

Untenable Demands

Luhmann and Ahn (2005) make several untenable demands of reasoners and of a theory of causal reasoning. In the previous section, we discussed their requirements that the reasoner (a) begins causal learning with knowledge of the true cause and (b) knows the boundaries of the learning context. In this section, we discuss their demands of a reasoning theory: (a) that it provide truth and (b) that its assumptions be individually validated.

Luhmann and Ahn (2005) criticize our theory for leading the reasoner to reach inaccurate estimates of causal power (and any estimate that is not 0 or 1 is inaccurate in their view). They portray a conception of hypothesis testing different from our own. Empirical truth is inevitably fallible. At best, reasoners can achieve only what the situation allows. In fact, they often fail to reach even the best possible truth—needless to say, they can hold erroneous assumptions. The claim that the power PC is a normative theory does not imply that reasoners will reach true estimates of causal power. The claim concerns the validity of the inference, rather than the truth of the conclusions.

Luhmann and Ahn (2005) also object to our proposal to test the assumptions in our theory by testing the theory's predictions; they instead advocate that the assumptions underlying our theory be individually validated. The issue they raise is not particular to our theory but concerns theory testing in general. The assumptions of a theory often involve unobservable entities that are not directly testable; that is the nature of theories. It is therefore often not

possible to test the individual assumptions underlying a theory, let alone ensure that they hold. Instead, a theory can be tested by its predictions on the basis of all of its assumptions operating in conjunction. Causal powers are theories, at the level both of particular inferred causal powers and of the power PC theory itself. Newell and Simon (1972), for example, hold a similar view regarding the testing of their theory:

> Our theory of human thinking and problem solving postulates that the human operates as an information processing system . . . we cannot reliably test the postulate by judging its intrinsic plausibility . . . there is little point in trying to judge directly whether a postulate is plausible. (p. 20)

We have evaluated our theory by testing its predictions. In view of our discussion on generalization to new contexts, we should report what our tests show about whether participants generalize to new contexts. Previous experiments testing causal power had the goal of discriminating between causal and purely covariational accounts of causal learning (Buehner & Cheng, 1997; Buehner, Cheng, & Clifford, 2003; Lober & Shanks, 2000; Novick & Cheng, 2004; Perales & Shanks, 2003). They therefore focused on testing predictions on the basis of the right-hand side of Equation 5 (or conjunctive versions of them). However, in those experiments in which participants were asked questions about a new context (in particular, a counterfactual context in which the alternative causes in the learning context were no longer present), there was indeed generalization (Buehner et al., 2003; Novick & Cheng, 2004), contrary to Luhmann and Ahn's (2005) hypothesis.[15]

Summary and Conclusion

Our commentators present theoretical and empirical arguments against our power PC theory (Cheng, 1997, 2000; Novick & Cheng, 2004). As we have shown in this article, however, people's flexible causal judgments across disparate situations can be coherently explained under our framework. This framework helps one avoid the assumption of erroneous constraints in the task of causal learning, such as disallowing legitimate input information and requiring knowledge of unobservable background causes. It also provides a clear criterion for evaluating the coherence of seemingly confusing concepts and the consistency of new types of causal judgments with preexisting theories. Contrary to White's (2005) arguments, judgments on releasing conditions, liabilities, and enablers; the influence of prevalence on causal attributions; and the acquisition of causal knowledge given only regularity information as data can all be explained by our theory. Contrary to Luhmann and Ahn's (2005) argument, a probabilistic causal power can result when no power PC assumption is violated, even for a reasoner who believes in causal determinism. Within our framework, it becomes transparent that the new type of confounding considered by Luhmann and Ahn leads to a logical contradiction. Their requirements that the reasoner (a) know the true cause and (b) restrict predictions in everyday causal inference to the learning context both lead to the paralysis of causal inference, denying the possibility of a task in which people evidently engage. We are grateful for the careful attention our commentators have given our work and for the opportunity to illustrate how a computational theory derived under a problem-solving perspective can guide the selection of coherent research questions.

[15] Luhmann and Ahn (2005) suggest that under the contexts presented in our cover stories testing simple causal power, participants are unlikely to compute simple causal power. For example, a participant in an experiment with a cover story about the influence of hypothetical drugs on patients would probably have prior knowledge that biological reactions would have complex causes involving other factors such as the patients' genetic makeup and environmental history. The fact that participants begin the experiment with such knowledge does not mean our framework is wrong or that our experimental results are invalid. Our framework describes the reasoning process that led up to that prior knowledge. (Obviously, reasoners need not have evaluated the relevant evidence themselves: The causal knowledge, possibly in an abstract form, can be culturally transmitted.) Thus, if participants thought, reasonably so, that contextual causal power was relevant, it would simply mean that simple causal power was evaluated and rejected, as would be consistent with our framework.

References

Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition, 54,* 299–352.

Buehner, M. J., & Cheng, P. W. (1997). Causal induction: The power PC theory versus the Rescorla-Wagner model. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 55–60). Hillsdale, NJ: Erlbaum.

Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariation to causation: A test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1119–1140.

Cartwright, N. (1989). *Nature's capacities and their measurement.* Oxford, England: Clarendon Press.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104,* 367–405.

Cheng, P. W. (2000). Causality in the mind: Estimating contextual and conjunctive causal power. In F. Keil & R. Wilson (Eds.), *Explanation and cognition* (pp. 227–253). Cambridge, England: MIT Press.

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology, 58,* 545–567.

Cheng, P. W., & Novick, L. R. (1991). Causes versus enabling conditions. *Cognition, 40,* 83–120.

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review, 99,* 365–382.

Griffiths, T. L., & Tenenbaum, J. B. (in press). Elemental causal induction. *Cognitive Psychology.*

Harré, R., & Madden, E. H. (1975). *Causal powers: A theory of natural necessity.* Oxford, England: Blackwell.

Holyoak, K. J. (2005). Analogy. In K. J. Holyoak & R. G. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 117–142). New York: Cambridge University Press.

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied, 79*(1, Whole No. 594).

Johnson, J. T., Boyd, K. R., & Magnani, P. S. (1994). Causal reasoning in the attribution of rare and common events. *Journal of Personality and Social Psychology, 66,* 229–242.

Johnson, J. T., Long, D. L., & Robinson, M. D. (2001). Is a cause conceptualized as a generative force? Evidence from a recognition memory paradigm. *Journal of Experimental Social Psychology, 37,* 398–412.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94,* 211–228.

Lien, Y., & Cheng, P. W. (2000). Distinguishing genuine from spurious causes: A coherence hypothesis. *Cognitive Psychology, 40,* 87–137.

Lober, K., & Shanks, D. R. (2000). Is causal induction based on causal power? Critique of Cheng (1997). *Psychological Review, 107,* 195–212.

Luhmann, C., & Ahn, W.-k. (2005). The meaning and computation of causal power: Comment on Cheng (1997) and Novick and Cheng (2004). *Psychological Review, 112,* 685–693.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.

Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal influence. *Psychological Review, 111,* 455–485.

Novick, L. R., Fratianne, A., & Cheng, P. W. (1992). Knowledge-based assumptions in causal attribution. *Social Cognition, 10,* 299–333.

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge, England: Cambridge University Press.

Perales, J. C., & Shanks, D. R. (2003). Normative and descriptive accounts of the influence of power and contingency on causal judgement. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 56*(A), 977–1007.

Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development, 47*(1, Serial No. 194).

Tenenbaum, J. B., & Griffiths, T. L. (2001). Structure learning in human causal induction. In T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems 13* (pp. 59–65). Cambridge, MA: MIT Press.

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General, 121,* 222–236.

White, P. A. (2000). Causal judgment from contingency information: The interpretation of factors common to all instances. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1083–1102.

White, P. A. (2004). Judgment of two causal candidates from contingency information: Effects of relative prevalence of the two causes. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 57*(A), 961–991.

White, P. A. (2005). The power PC theory and causal powers: Comment on Cheng (1997) and Novick and Cheng (2004). *Psychological Review, 112,* 675–684.

Wu, M., & Cheng, P. W. (1999). Why causation need not follow from statistical association: Boundary conditions for the evaluation of generative and preventive causal powers. *Psychological Science, 10,* 92–97.

Received December 5, 2004
Revision received March 14, 2005
Accepted April 1, 2005 ■

Postscript

Patricia W. Cheng
University of California, Los Angeles

Laura R. Novick
Vanderbilt University

We briefly reply to the main points in White's (2005) and Luhmann and Ahn's (2005) *Postscripts*. Regarding the ingestion-of-medicine example, the prior knowledge that White uses about the other three cells in the contingency tables (see Figure 2 of our reply to the comments; Cheng & Novick, 2005) may not be apparent as input until one manipulates the outcomes in these cells. As our figure illustrates, different outcomes in these cells yield different causal conclusions. As for whether White's (2000) confounded results support his model of causes and enablers, a replication of the relevant condition without the problems noted earlier will provide the answer. Regarding the possibility of direct causal knowledge, causal inference involving haptic input will share a common core of constraints with causal inference involving other modalities—a core that includes regularity information. The immediacy and effortlessness of the perception of causality does not imply that no computation is involved, just as the immediacy and effortlessness of visual perception does not indicate that visual perception involves no computation, as any attempt to build a machine to perform those tasks will show.

Luhmann and Ahn's (2005) argument that q_{citrus}, the causal power of citrus fruit with respect to beetle repelling in our example, should be 0 despite q_{orange} being 1 is yet another demonstration of the incoherence of their framework. As they show in their Footnote 2, the value of 0 for q_{citrus} is what would be required for consistency between the two sides of their Equation 1 (from Cheng, 1997). But, $q_{\text{citrus}} = 0$ is incompatible with $q_{\text{orange}} = 1$ if oranges exist (as they do in our example), given the meaning of causal power. As noted earlier, q_x, the causal power of candidate cause x with respect to effect e, is defined as "the probability with which x produces e when x is present" (Cheng, 1997, p. 372). If oranges repel beetles, then when citrus fruits are present, the oranges among them repel beetles, and therefore citrus fruits evidently repel beetles with a nonzero probability. In summary, the 0 value that Luhmann and Ahn's framework requires for consistency in the equation is in fact logically impossible. Contrary to their argument, the value of 0.5 for q_{citrus} we used in our demonstration is not confounded. Our demonstration shows that if the values for q_{citrus} and q_{orange}, which are alternative hypotheses, are respectively obtained according to our theory, then treating alternative hypotheses as alternative causes would lead to counting the effect of the same token of the candidate cause multiple times (*e* due to orange O, a particular orange, as an orange and again as a citrus fruit).

Luhmann and Ahn (2005) argue for a framework of causal learning in which no causal learning can take place. They are unable to disagree with the resulting paralysis of causal inference noted earlier. They concur, for example, that the causal power of citrus peels in our beetle repellent example is unknowable by a reasoner unless he or she is omniscient. An overriding question is why their position would be worth considering. In keeping with their failure to treat causal learning as a problem to be solved, they argue for an answer that is not a solution to any problem.

The rest of Luhmann and Ahn's (2005) *Postscript* commits errors already noted earlier. First, they confuse the truth of a conclusion with the validity of an inference. They fault our theory for giving inaccurate causal powers when the proportion of oranges in citrus fruits varies across contexts. But, no theory of reasoning can guarantee true conclusions. The disconfirmation of a causal power inferred in a new context, rather than being detrimental to our theory, is in fact helpful to the construction of a more accurate picture of the causal world, as Cheng (2000) noted. In our example, if the reasoner notices that there is variation in q_{citrus} across contexts and that q_{citrus} correlates with the proportion of